ONLINE APPENDIX

Appendix A. Correlates of Villagization

Our empirical strategy exploits variation in villagization intensity across space, as well as variation in age cohorts that were differentially exposed to the *Ujamaa* education reform. Variation in villagization intensity across space can largely be explained by the fact that its implementation was left to district officials, who enforced the policy with different degrees of conviction (McHenry, 1979). This variation may not be random across space (nor does it need to be for our empirical strategy to be valid, as we explain in Section 4). Nevertheless, we control for the most important potential correlates of villagization intensity, as identified in historical accounts. These include: the pre-*Ujamaa* primary school enrollment rate, local government capacity, ethnolinguistic fractionalization, geographic characteristics, availability of public health infrastructure, weather, the degree of urbanization, initial national identity, population density, levels of economic development, agricultural production, distance to Uganda and the presence of colonial missions and roads (Ergas, 1980; Hydén, 1980). Table A.11 documents the pairwise correlations of these variables with villagization (conditional on zone fixed effects, as in our baseline specification). We use newly digitized district-level variables from official government statistics (Jensen, 1968) based on the 1967 population census, the last census before the villagization policy was implemented, and various other sources (see Table A.2 for variable definitions and data sources).

Pre-policy Access to Public Services

The data shows a positive correlation between our measure of villagization (share of the rural district population living in planned villages in 1978) and the pre-policy primary school enrollment rate. *A priori*, we might have expected the government to target areas with low levels of previous access to schooling given the policy's goal to build national identity through public education. However, this positive correlation could be explained by the fact that the government was more successful in implementing the policy in areas that already had existing (schooling) infrastructure. The government used access to schools and other public services such as dispensaries and piped water supply to encourage people to move to *Ujamaa* villages (McHenry, 1979; Boesen et al., 1977; Scott, 1998). Settlements that already featured some of this infrastructure thus made for easier targets. In any case, primary school enrollment levels were low (7%) prior to villagization.

In addition, we test whether there is a relationship between the intensity of villagization and existing public health infrastructure, proxied by the number of hospital beds per hundred thousand inhabitants in 1967. We find no significant correlation.

Pre-policy District Capacity

Given that the policy was implemented primarily by district officials (McHenry, 1979), we might expect districts with higher local state capacity to implement the policy more intensely – both in terms of resettling the local population and implementing the concurrent education reform. The central government did not have sufficient funds to build all of the new schools required to meet the universal primary education targets laid out in the policy, and thus promoted "self-help" efforts by district and village officials (Stabler, 1979). We proxy for local state capacity using district government revenue per capita. Contrary to what we might expect given the historical literature, we find a negative but close to zero correlation between pre-*Ujamaa* government revenue and villagization.

However, we do find that villagization intensity is weakly positively correlated with the pre-*Ujamaa* degree of urbanization (1 – share of the rural district population in 1967). While the policy was rural in nature, the government may have been more successful at consolidating the population in districts with existing settlements, as noted above. In any case, urbanization levels were very low (3% of the country's population according to the 1967 population census) prior to villagization.

The number of missions per capita and the length of roads divided by district area are also not significantly correlated with villagization. The former is negatively correlated, pointing to the government's efforts to reach the most peripheral areas, while the latter is positive. Areas with higher road density might have been relatively easier to reach, favoring villagization, even if only slightly.

Pre-policy Ethnolinguistic Fractionalization

We next look at the relationship between villagization and ethnolinguistic fractionalization (ELF) in 1967. McHenry (1979) argues that ethnic diversity hindered the population's willingness to comply with the villagization policy, which involved potentially co-inhabiting villages with members of different ethnic groups. Consistent with this argument, we find that the intensity of villagization was lower in districts with high pre-*Ujamaa* ELF.

Distance to Dar es Salaam (Central Government Influence)

We next consider the correlation between the distance to the capital city, Dar es Salaam[51], as a measure of central government influence, and the intensity of villagization. Given limited control of the hinterland by the state (Herbst, 2014), we might expect that the policy was particularly targeted to the remote parts of the country to bring those areas into the state's

[51] Dar es Salaam was not only the seat of government in the pre-policy era but also the country's main administrative and commercial center. In 1973, the government made the decision to move the national capital to Dodoma within 10 years (Hayuma, 1980).

orbit. Indeed, Scott (1998) also notes that sites for new village locations were chosen not by any economic logic, but by finding "blank spots" on the map where settlers might be relocated citing Boesen et al. (1977). Consistent with expectations based on these historical accounts, we find a positive relationship between distance to Dar es Salaam and villagization intensity.

Geography

Historical accounts indicate that the most important criterion for the location of Ujamaa villages was convenient access (Hydén, 1980). This implies that geographic characteristics such as topology or climatic zones may have also influenced the degree of villagization. For example, more rugged or remote terrain may have been harder to consolidate. However, we find no significant correlations between several geographic characteristics (altitude, slope, latitude and longitude) and villagization.

Weather Shocks

Several regions in Tanzania experienced severe droughts during the height of the resettlement period (1974–1976). Hydén (1980) argues that from the point of view of the state, the drought served as a "blessing in disguise [that] facilitated the movement of people" (p. 146). The drought resulted in a major food shortage (Ergas, 1980). The government also provisioned drought relief to incentivize people to move into planned villages. Indeed, in a study on the motives behind migration to villages, Bakula (1971, in Sitari (1983)) finds that 80% of migrants say main reason they moved was hope of increased monetary income. Thus, we check whether unexpected weather shocks in the form of droughts affected villagization. We find a positive correlation between drought occurrence in 1975 and 1976, and villagization.[52]

Pre-policy Economic Development

Scott (1998) notes that some regions (in particular, West Lake and Kilimanjaro) saw very little impact of *Ujamaa*. These regions were spared for three reasons: (1) farmers were already living in populous villages; (2) their undisturbed productivity of cash crops was vital for state revenue; and (3) groups residing in these areas were over-represented in the bureaucratic elite. Ergas (1980) and Sitari (1983) echo this, noting the poorest response to villagization in densely settled, prosperous territories. These areas were home to prosperous farmers, and were places where cash crops were produced in large amounts. Indeed, according

[52] We measure droughts using the standard deviation in yearly rainfall from a district's long-term mean rainfall (1960–2010) during the planting season (March to May), censored at 0 (following Dell et al. (2014)). Rainfall is calculated based on readings from 107 weather stations across Tanzania and aggregated to the district level using Kriging (spatial interpolation).

to Barker (1974), the number and population of *Ujamaa* villages appears to vary inversely with importance of cash crops in region. These historical accounts would lead us to expect a negative correlation between villagization intensity and pre-policy economic development. However, it seems that villagization was *stronger* in areas with higher agricultural production and per capita GDP before *Ujamaa*, conditional on zone fixed effects, albeit the correlations are not significant.

Distance to Ugandan Border

Villagization occurred during a period that spanned war with Uganda. In response to Idi Amin's annexation of Kagera region in northwestern Tanzania in November 1978, Nyerere launched a counter-offensive that ultimately removed Amin from power in April 1979 (Roberts, 2014). Although the official loss to Tanzania in terms of human life was relatively low (373 soldiers killed, of whom 96 died due to enemy fire), the ultimate cost of the intervention and occupation of Uganda was estimated to be more than $500 million (Roberts, 2014). We find that villagization was stronger the farther away from the border with Uganda, with a relatively large but insignificant correlation coefficient.

Pre-policy Feelings of National Identity

Lastly, in Figure A.8, we report the association between villagization in 1978 and feelings of national identity in 1967 (i.e, prior to the *Ujamaa* policy) at the region level.[53] The data on national identity in 1967 is based on a nationwide survey of 3,000 secondary students (Prewitt et al., 1970) and is similar in wording to the measure of national identity from the contemporary Afrobarometer data used in the paper. We find a negative relationship between villagization and pre-*Ujamaa* feelings of national identity. This suggests that, in line with the government's stated intention of building national identity using villagization, the policy was aimed at regions that historically had lower levels of identification with the nation.

Appendix B. Migration

Here we conduct two exercises to show that selective migration is unlikely to explain most of the positive relationship between the *Ujamaa* policy and national identity we observe. Recall that only inter-district migration potentially threatens the validity of our estimates.

The Afrobarometer data do not contain information on respondents' migration. However, we can calculate inter-district migration rates, during the period after *Ujamaa* and before

[53] The 1967 survey by Prewitt et al. (1970) does not contain district-level identifiers. We thank Ted Miguel and Tina Green for help with accessing the data.

the Afrobarometer survey took place, using the Tanzania National Panel Survey 2008 data.[54] That is, we can calculate the in- and out-migration rate for each origin-/destination-district-combination. We use the migration matrix constructed this way for the robustness checks related to selective migration.

As a first exercise, in Table 3 column (5) we drop 14 out of 52 districts that are in the top deciles in terms of either in- or out-migration rates. If (selective) migration were driving our results, we should see a clear drop in our main coefficient when we exclude such high-migration districts. Reassuringly, even though we lose some statistical power due to the reduced sample size, the magnitude of the coefficient with this reduced sample is similar to the coefficient for our baseline specification with the full sample.

As a second exercise, we quantify the potential measurement error stemming from selective migration to provide bounds on our main coefficient under different assumptions about the extent of selective migration. We begin by noting that migration may or may not lead to measurement error in a respondent's villagization measure: for example, if a respondent in the treated cohort is at the time of survey (2005-2008) in a district with a villagization measure of 0.5, but was in a district with a villagization measure of 1 during the villagization period (1970–1981), then the measurement error in her treatment amounts to -0.5. However, if she migrated from a district with a villagization measure of 1 to another district with a villagization measure of 1, the measurement error is 0 and does not bias our estimates.

The expected measurement error in the villagization measure among respondents in a district thus depends on two factors: 1) the difference in villagization between the receiving district and the sending district, and 2) the migration rates between each district-pair. To estimate the expected measurement error in our sample, we proceed as follows. We first calculate the measurement error for each district-pair (receiving and sending) as the product of the difference between the two districts' villagization measures and the rate of migration from the sending to the receiving district:[55]

$$error_{ds} = (V_d - V_s) \cdot m_{ds} \tag{A 1}$$

where V_d and V_s denote villagization in receiving and sending districts d and s, and $m_{d,s}$ denotes the migration rate from district s to district d. Migration from districts with higher

[54] The Tanzania National Panel Survey data contains variables on each respondent's current district, previous district, and year of migration (if migrated). We calculate migration rates based on migration during 1970 and 2004 because we are concerned about migration caused by the *Ujamaa* policy in 1970-1981 before the Afrobarometer survey (2005-2008). We assume that if the respondent migrated between 1970 and 2004, the origin district is the district where she resided during the villagization period.

[55] For example, assume that district A has a villagization measure of 0.8 and district B has a villagization measure of 0.6. Further assume that 5% of current residents in district A migrated from district B. Then, the expected measurement error in the villagization measure among the current residents of district A stemming from migration from district B equals $(0.8 - 0.6) \cdot 0.05 = 0.01$.

villagization thus leads to negative measurement error in the villagization measure, while migration from districts with lower villagization leads to positive measurement error. The larger the migration rate from a district, the larger the potential measurement error due to migration from that district.

Next, we make an extreme assumption that would lead to selective migration working against our main finding: assume that residents in the current district with the national identity outcome equal to y (ranging from 0 to 1) came from a given district with lower villagization with probability $m_{ds} \cdot y_{id}$ and from a given district with higher villagization with probability $m_{ds} \cdot (1 - y_{id})$. That is, residents with higher (lower) national identity were more likely to migrate from districts with lower (higher) villagization. We adjust the villagization measures of the respondents in our main sample by the resulting expected error after summing over sending districts:[56]

$$\tilde{V}_{id} = V_d - \sum_{s=1}^{S} error_{ds} \cdot \mathbb{1}\left[error_{ds} > 0\right] \cdot y_{id} - \sum_{s=1}^{S} error_{ds} \cdot \mathbb{1}\left[error_{ds} < 0\right] \cdot (1 - y_{id}) \quad \text{(A 2)}$$

where $\tilde{V}_{id}$ denotes individual i in district d's "error-adjusted" villagization measure and y_{id} denotes national identity (ranging from 0 to 1).

Finally, we re-estimate our main specification using this adjusted villagization measure:

$$y_{idzt} = \gamma(\tilde{V}_{id} \cdot treatedcohort_t) + (\tilde{\mathbf{X}}'_{dz} \cdot treatedcohort_t)\mathbf{\Gamma} + \alpha_{dz} + \delta_{zt} + \epsilon_{idzt} \quad \text{(A 3)}$$

The coefficient estimate of γ then provides a lower bound to our baseline coefficient estimate of β in the presence of selective migration to such an extent. Table 3 column (6) displays the result: the estimated coefficient equals 0.142 ($p < 0.05$). As a reminder, our baseline coefficient estimate equals 0.226. That is, at most 37% of our baseline coefficient estimate can be explained by such selective migration. However, this case is very unlikely. It effectively assumes that individuals that migrated to low-villagization districts after villagization ended identify with their ethnic group, and that individuals that migrated to high-villagization districts after villagization ended identify with Tanzania as a whole. Alternatively, assume that only some of those with low national identity systematically migrated to low-villagization districts after the end of the *Ujamaa* period, whereas there was no selective migration among those with high national identity (that is, they did not systematically go to districts with higher villagization after villagization ended). In this case, the adjusted coefficient estimate, reported in Table 3 column (7), equals 0.164 ($p < 0.05$). That is, under this more realistic assumption, at most 27% of our main result can be explained by selective migration of this nature. In sum, under reasonable assumptions, it is unlikely that selective migration explains a large fraction of our main result.

[56] We adjust the 1967 district primary school enrollment rate control variable analogously.

Appendix C. Additional Robustness Checks

Here we discuss robustness checks related to alternative choices of regression weights, sample and functional form, and address concerns related to social desirability bias.

Sample weighting. In our preferred specification, we use survey weights from Afrobarometer to make our sample nationally representative. Table 3 column (2) shows the unweighted coefficients. The estimates from the weighted and unweighted regressions are qualitatively similar in size and significance.

Sample choice. Since villagization was a rural policy, those living in rural areas during the *Ujamaa* period should be affected by the treatment (the "compliers"). Table 3 column (4) reports our coefficient estimates for the sample of respondents living in a rural area at the time of the survey. As expected, we find that the effect of *Ujamaa* on national identity is larger for this sub-sample (as they are more likely to be compliers).

Functional form. In Table 3 columns (10) and (11), we report results for ordered probit and ordered logit specifications. The results are consistent with our baseline result, which is based on the linear probability model. See the table notes of Table 3 for details.

Social desirability bias. Our main outcome measure, whether the respondent reports identifying more with her ethnic group or with the nation as a whole, may suffer from social desirability bias. The treated respondents were taught the importance of the nation and a sense of duty towards the state as part of the *Ujamaa* policy. Hence, treated respondents may believe that they are expected to answer in a certain way because they may believe that the surveyor is a representative of the national government. This could bias our coefficient of interest upwards. Afrobarometer spends considerable effort to mitigate such concerns by hiring independent surveyors unaffiliated with the government and from outside the survey districts. Nevertheless, almost 60% of the respondents in our sample report thinking that the surveyors were sent by a government institution.

Social desirability bias is unlikely to explain our results for at least two reasons. First, we directly control for whether a respondent believes that the surveyor was sent by a government institution. Our estimates are unchanged when including this control variable (see Table 3 column (3)). Second, our results on inter-ethnic marriage (discussed below), which we view as a "revealed preference" measure of identity that should not suffer from potential reporting bias, are very consistent with the estimated effect on the main measure of national identity from the Afrobarometer.

Appendix D. The End of *Ujamaa*

A question is whether the external developments that explain the end of *Ujamaa*, including a global economic crisis and the invasion by Uganda (see the Background section for a summary), had cohort-specific effects that could explain our main result. We address this question in two ways.

First, we test whether important factors leading to the end of *Ujamaa*, including economic factors and the war with Uganda, could explain our results. In brief, we do not find any empirical support for such an alternative explanation. Economic factors are unlikely to explain our cohort-specific results given the results in Table A.8 discussed in the paper. In addition, Table A.5 column (8) shows our estimate of the effect of *Ujamaa* on national identity is unaffected by the inclusion of distance to the Ugandan border interacted with the cohort dummy, which proxies for cohort-specific exposure to the Ugandan invasion.

Second, to assess whether the end of *Ujamaa* had cohort-specific effects, we exploit variation across cohorts that were all not exposed to our treatment directly and are thus comparable but that differ in their exposure to the end of *Ujamaa*. Specifically, we compare the cohort born in 1948–1959 (i.e., those too old to attend primary school during *Ujamaa* but who experienced the end of the villagization program) to a younger "control" cohort born in 1976–1987 (i.e., cohorts entering primary school only after the end of the villagization program) in more or less intensively villagized districts.[57] Table 3 column (12) shows that we do not find any economically or statistically significant differences in national identity between these cohorts, and thus conclude that the end of *Ujamaa* did not have cohort-specific effects.

[57] Note, one could also compare the outcomes of the treated cohorts (i.e., those in highly villagized districts during the villagization program) with younger cohorts (i.e., cohorts entering primary schooling age after the end of the villagization program). However, this comparison would not allow for separate identification of the effect of exposure to the end of *Ujamaa* and *Ujamaa* itself.

Appendix E. Additional Tables

Table A.1: Timeline

Year	Event	Details
1961	Independence	Tanganyika (Tanzania Mainland) becomes independent with Julius Nyerere as prime minister.
1963	One-party state declared	Nyerere makes TANU the sole national party in the interest of the whole nation and economic development (Nyerere, 1963).
1967	National vision outlined	*Arusha Declaration* as basis for national ideology; statements by party and prime minister announcing plans for nation-building policy and *Ujamaa* villages.
1967	*Education for Self-Reliance*	Formed basis of *Ujamaa* education reforms; Swahili made language of instruction in public primary schools.
1967	1967 Population Census	First population census after independence conducted; administrative districts defined.
1968	Roll-out of new education policies	Political education added as a new subject to primary school curriculum (grades 4 to 7); new syllabi and textbooks emphasizing national identity introduced in social studies, history, geography and civics.
1969	**Presidential Circular No. 1**	**Mandatory villagization begins.**
1975	*Villages &* Ujamaa *Villages Act*	Villagization formalized in the law.
1978	1978 Population Census	First census after villagization was declared concluded.
1982	**Repeal of *Villages* & Ujamaa *Villages Act***	**Villagization period ends.**
1985	Nyerere steps down	Ali Hassan Mwinyi replaces Nyerere as prime minister, abandons many of Nyerere's policies and enacts reforms.

Notes: This table provides a timeline of activities by Nyerere and the Tanzanian state after Independence and during the *Ujamaa* period. See sources described in Section 2 for additional details.

Table A.2: District-level Variables: Sources and Definitions

Variable	Source	Details and Coding
Villagization 1978	Population Census	Total rural district population living in registered villages (as opposed to unregistered villages, scattered and migratory population) divided by total district population, excluding the institutional population, as classified by the 1978 population census.
Primary school enrollment rate in 1967	Jensen (1968)	Number of children enrolled in primary schools per one thousand inhabitants according to official government statistics (Jensen, 1968).
Distance to Dar es Salaam	UC Davis DataLab, HDX	Distance of the district centroid to Tanzania's capital city in kilometers.
District revenue per capita in 1966	Jensen (1968)	Total district revenue per capita in thousands of Tanzanian Shillings according to official government statistics (Jensen, 1968).
Ethnolinguistic fractionalization in 1967	Population Census	Calculated as 1 minus the Herfindahl concentration index of ethnolinguistic group shares in each district given by the 1967 population census data.
Centroid latitude / longitude	UC Davis DataLab, HDX	Latitude / longitude of the district centroid in degrees.
Average altitude	UC Davis DataLab, HDX	Mean district altitude in meters, calculated from gridded topology data.
Average slope	UC Davis DataLab, HDX	Mean slope, calculated from gridded topology data.
Hospital beds in 1967	Jensen (1968)	Number of hospital beds per 100,000 inhabitants according to official government statistics (Jensen, 1968).
Drought in 1974-1976	Tanzania Meteo	Average amount of yearly rainfall per district during the planting season (March to May), in terms of standard deviations from the long-run district mean (1960-2008), censored at 0. Data constructed from readings of 107 weather stations using Kriging (spatial interpolation).
In- / out- migration rate 1970-2004	TNPS	Fraction of district population that moved from/to a different district between 1970 and 2004. Calculated from individual-level data on household heads in the TNPS Round 1 (2008).
National identity in 1967	Prewitt et al. (1970)	Based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 (Prewitt et al., 1970). The variable is coded as 1 if a respondent feels that the nation is more important than the tribe and 0 otherwise.
Share urban 1967	Jensen (1968)	Share of total district population living in urban wards according to official government statistics (Jensen, 1968).
Population density in 1967	Jensen (1968)	Total district population in 1967, divided by 1967 district area (square miles) (Jensen, 1968).
GDP per capita in 1967	Jensen (1968)	Total district GDP in 1967, divided by district population in 1967 (Jensen, 1968).
Tot. market agricultural production per capita in 1967	Jensen (1968)	Total district market agricultural production in 1967, divided by district population in 1967 (Jensen, 1968).
Distance to Uganda	UC Davis DataLab, HDX	Distance to the closest border with Uganda, in km.
Number of missions per capita in 1968	Hedde-von Westernhagen and Becker (2022)	Number of colonial missions in the district in 1968, divided by district population in 1967 (Hedde-von Westernhagen and Becker, 2022).
Total length of roads over district area in 1968	Jedwab and Storeygard (2022)	Total length of roads in the district in 1968 (in km), divided by district area, in square km. (Jedwab and Storeygard, 2022).
CCM support in elections	Electoral data	Product of turnout and CCM presidential vote share from the 2000 and 2005 presidential elections in Tanzania (data from Carlitz (2017)), aggregated at the level of 1967 districts averaging over constituencies.

Notes: This table provides the key district-level variables, their sources and coding definitions used in our empirical analysis. There are 52 districts in our main sample. Section 3 provides additional discussion.

Table A.3: Individual-level Variables: Sources and Definitions

Variable	**Source**	**Details and Coding**
National identity	Afrobarometer	"Let us suppose that you had to choose between being a [nationality] and being a [respondent's identity group]. Which of the following statements best expresses your feelings?" 0 = ethnic only, 1/4 = ethnic more than national, 1/2 = ethnic and national equal, 3/4 = national more than ethnic, 1 = national only.
School cohort	Afrobarometer	Constructed from birth year of respondent plus 10 years (typical age of 4th grader). "Treated cohort" = born 1960-1971, "control cohort" = born 1948-1959.
Believe surveyor sent by government	Afrobarometer	"Who do you think sent us to do this interview?" 1 = government or other public institution, 0 = otherwise.
Urban	Afrobarometer	"Urban or Rural Primary Sampling Unit" 0 = rural, 1 = urban.
Completed school	Afrobarometer / IPUMS	"What is the highest level of education you have completed?" 1 = completed formal primary schooling, 0 = completed no formal schooling.
Respect authority	Afrobarometer	"Let's talk for a moment about the kind of society we would like to have in this country. Which of the following statements is closest to your view? Choose Statement A or Statement B. A: As citizens, we should be more active in questioning the actions of our leaders. B: In our country these days, we should show more respect for authority." 0 = agree very strongly with A, 1/4 = agree with A, 1/2 = agree with neither, 3/4 = agree with B, 1 = agree very strongly with B.
Support one-party rule	Afrobarometer	"There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." 0 = strongly disapprove, 1/4 = disapprove, 2/4 = neither approve nor disapprove, 3/4 = approve, 1 = strongly approve.
See government as parent	Afrobarometer	"Lets talk for a moment about the kind of society you would like to have in this country. Which of the following statements is closest to your view? Choose Statement 1 or Statement 2.Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government."0 = agree very strongly with 2, 1/4 = agree with 2, 1/2 = agree with neither, 3/4= agree with 1, 1= agree very strongly with 1.
Trust government newspapers / TV / radio	Afrobarometer	"How much do you trust each of the following, or haven't you heard enough about them to say: Government / independent newspapers / TV / radio?" 0 = not at all, 1/3 = just a little, 2/3 = somewhat, 1 = a lot. For each medium, the reported trust for the independent medium is subtracted from the reported trust for the Government one.
Took action to hold government accountable	Afrobarometer	"Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: Got together with others to raise an issue? Attended a demonstration or protest march? " 0 = if has never done any of these things, 1 = has done at least one of these things at least once.
Trust own / other ethnic group(s)	Afrobarometer	"How much do you trust each of the following types of people: People from your own ethnic group? / Tanzanians from other ethnic groups?" 0 = not at all, 1/3 = Just a little, 2/3 = I trust them somewhat, 1 = I trust them a lot
Cash income job	Afrobarometer	"Do you have a job that pays a cash income? Is it full-time or part-time?" 0 = no, 1/2 = yes, part time, 1 = yes, full time.
Wealth index	Afrobarometer	First principal component of the questions asking about the frequency of the household not having enough food, clean water, medicines, cooking fuel, cash income, and about ownership of a radio, a tv and a motorbike.
Support free schooling	Afrobarometer	"Which of the following statements is closest to your view? A: It is better to have free schooling for our children, even if the quality of education is low. B: It is better to raise educational standards, even if we have to pay school fees." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.
Agree women should have equal rights	Afrobarometer	"Which of the following statements is closest to your view? A: In our country, women should have equal rights and receive the same treatment as men do. B: Women have always been subject to traditional laws and customs, and should remain so." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.

Continued on next page

Table A.3 – continued from previous page

Variable	Source	Details and Coding
Agree women should have equal election chances	Afrobarometer	"Which of the following statements is closest to your view? A: Women should have the same chance of being elected to political office as men. B: Men make better political leaders than women, and should be elected rather than women." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.
Support CCM	Afrobarometer	"Do you feel close to any particular political party?" 1 = yes, Chama cha Mapinduzi (CCM), 0 = yes, party different than CCM.
Kiswahili Literate	IPUMS	"Does individual know how to read and write in Kiswahili?" 1 = yes, 0 = no.
Ann. real household consumption, per adult	TNPS	Total annual real household consumption per adult equivalent.
Ann. nom. furnishings and household expenditures	TNPS	Total annual nominal household expenditures on furnishings and household items.
Main occupation	TNPS	Dummies for different occupations that equal 1 if household head's occupation is in government (including parastatal) / private sector / agriculture / self-employment, and 0 otherwise.
Married within ethnic group	DHS	= 1 if respondent's spouse shares self-reported ethnic group, 0 otherwise. Includes married couples only.
Married	DHS	= 1 if respondent is currently married, 0 otherwise.

Notes: This table provides the key individual-level variables, their sources and coding definitions used in our empirical analysis. Section 3 provides additional discussion.

Table A.4: The Effect of *Ujamaa* on National Identity, All Cohorts

	Dependent Variable: National vs. Ethnic Identity							
	no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with					
			distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Birth Year 1952-1955	-0.006	0.014	-0.005	-0.047	-0.013	-0.046	-0.083	0.091
	(0.158)	(0.168)	(0.170)	(0.172)	(0.169)	(0.204)	(0.227)	(0.192)
Villagization × Birth Year 1956-1959	-0.034	0.008	-0.016	0.011	-0.037	-0.047	0.006	-0.012
	(0.075)	(0.076)	(0.085)	(0.063)	(0.089)	(0.097)	(0.079)	(0.082)
Villagization × Birth Year 1960-1963	0.135	0.170	0.145	0.126	0.099	0.077	0.161	0.110
	(0.123)	(0.149)	(0.153)	(0.128)	(0.145)	(0.131)	(0.141)	(0.127)
Villagization × Birth Year 1964-1967	0.120	0.223*	0.182	0.195*	0.167	0.138	0.231*	0.227*
	(0.098)	(0.122)	(0.115)	(0.106)	(0.101)	(0.114)	(0.124)	(0.133)
Villagization × Birth Year 1968-1971	0.176	0.284**	0.260**	0.255***	0.217**	0.206	0.287**	0.287**
	(0.109)	(0.107)	(0.112)	(0.094)	(0.105)	(0.126)	(0.110)	(0.123)
Villagization × Birth Year 1972-1975	0.062	0.139	0.106	0.108	0.083	0.007	0.136	0.148
	(0.090)	(0.097)	(0.112)	(0.089)	(0.098)	(0.113)	(0.093)	(0.103)
Villagization × Birth Year 1976-1979	0.095	0.159	0.122	0.128	0.086	0.075	0.168*	0.164
	(0.090)	(0.108)	(0.114)	(0.119)	(0.124)	(0.145)	(0.098)	(0.123)
Villagization × Birth Year 1980-1983	0.016	0.060	0.044	0.027	0.029	-0.002	0.057	0.064
	(0.081)	(0.103)	(0.115)	(0.085)	(0.093)	(0.099)	(0.102)	(0.090)
Villagization × Birth Year 1984-1987	0.020	0.032	0.026	0.011	-0.029	-0.076	0.009	-0.000
	(0.099)	(0.130)	(0.137)	(0.125)	(0.134)	(0.139)	(0.107)	(0.128)
Observations	1,797	1,797	1,797	1,797	1,797	1,797	1,797	1,797
Number of clusters	52	52	52	52	52	52	52	52
R-squared	0.087	0.111	0.116	0.118	0.115	0.135	0.122	0.131
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to birth cohort t. All columns regress the measure of an individual's current national identity on the interactions between that individual's district-level measure of historical villagization and dummies that indicate the individual's cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummies, survey year fixed effects, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). The reference group is the cohort born in 1948-1951. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort fixed effects, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall during the planting season in 1974, 1975 and 1976 (main period of forced relocation), censored at 0 to capture droughts. The weather data is from UC Davis DataLab, HDX, and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: The Effect of *Ujamaa* on National Identity, Additional Controls

	Dependent Variable: National vs. Ethnic Identity									
	no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with							
			national identity in 1967	share urban	population density	GDP	agricultural production	distance to Uganda border	colonial missions	roads length
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Villagization × Treated Cohort	0.165**	0.226***	0.173**	0.229***	0.232***	0.226***	0.236***	0.234***	0.227***	0.227***
	(0.079)	(0.081)	(0.075)	(0.085)	(0.084)	(0.079)	(0.076)	(0.085)	(0.081)	(0.080)
Observations	849	849	785	849	849	849	849	849	849	849
Number of clusters	52	52	48	52	52	52	52	52	52	52
R-squared	0.121	0.139	0.137	0.139	0.139	0.139	0.140	0.139	0.139	0.139
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level. Added controls in columns (3) to (10) are, respectively, national identity in 1967 based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 by Prewitt et al. (1970), urbanization, population density, GDP, the value of agricultural production (all from Jensen (1968)), distance to the border with Uganda, the number of colonial missions per capita (Hedde-von Westernhagen and Becker, 2022), and the total length of roads in the district divided by district area in 1968 (Jedwab and Storeygard, 2022). The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: The Effect of *Ujamaa* on Primary Education and Literacy in Kiswahili

Dep. Variable:	Primary School (1)	Kiswahili Literate (2)
Villagization × Treated Cohort	0.009*	0.008*
	(0.005)	(0.005)
Observations	562,722	653,272
Number of clusters	52	52
R-squared	0.152	0.085
District FE	✓	✓
Cohort FE	✓	✓
Zone-Cohort FE	✓	✓

Notes: The unit of observation is an individual IPUMS 1988 respondent i in district d belonging to school cohort t. The dependent variable in column (1) is a dummy for whether the respondent completed primary school and in column (2) a dummy for whether the respondent is literate in Kiswahili. All columns regress the dependent variable on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Respondents with partial primary schooling are excluded from the sample in column (1) and those who are still attending school are excluded from the sample in both columns (1) and (2). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by IPUMS. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: The Effect of *Ujamaa* on Occupation

	Dependent Variable: Occupation			
	government and parastatal (1)	private sector (2)	agriculture (3)	self-employed (4)
Villagization × Treated Cohort	-0.004	0.005	0.024	-0.016
	(0.005)	(0.006)	(0.021)	(0.020)
Observations	1,599	1,599	1,599	1,599
Number of clusters	52	52	52	52
R-squared	0.053	0.050	0.142	0.111
District FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓

Notes: The unit of observation is an individual Tanzania National Panel Survey Round 1 (2008) respondent i in district d belonging to school cohort t. All columns regress a dummy for the occupation listed in the column head on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the household survey weights provided by the data. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: The Effect of *Ujamaa* on Economic Outcomes

Dep. Variable:	HH Consumption	HH Expenditures	Cash Income Job	Wealth Index	National Identity	
	(1)	(2)	(3)	(4)	(5)	(6)
Villagization × Treated Cohort	-0.184***	-0.034	-0.120**	-0.099		
	(0.057)	(0.033)	(0.050)	(0.072)		
Cash Income Job					0.032	
					(0.037)	
Wealth Index						0.073*
						(0.041)
Observations	958	958	868	859	849	841
Number of clusters	52	52	52	52	52	52
R-squared	0.137	0.095	0.187	0.178	0.134	0.138
District FE	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓
Source	TNPS	TNPS	Afrobarometer	Afrobarometer	Afrobarometer	Afrobarometer

Notes: The unit of observation is an individual Tanzania National Panel Survey Round 1 (2008) respondent, in columns (1) and (2), and an Afrobarometer respondent, in columns (3) to (6), i in district d belonging to school cohort t. The sample in columns (1) and (2) is restricted to household heads. Columns (1) to (4) regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Column (5) regresses the national identity measure on having a job that pays cash income and column (6) on a wealth index, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. Household consumption is total annual real household consumption per adult equivalent. Household expenditures are total annual nominal household expenditures on furnishings and household items. Cash Income Job is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a three-level scale; it has been recoded from 0 to 1 in 1/2 increments (1 = respondent has a full time job that pays cash income; 1/2 = respondent has a part time job that pays cash income; 0 = respondent does not have a job that pays cash income). The wealth index is from Afrobarometer rounds 3 and 4 (2005-2008), and is obtained from the first principal component of the questions asking about the frequency of the household not having enough food, clean water, medicines, cooking fuel, cash income, and about ownership of a radio, a tv and a motorbike. These questions are present in both rounds 3 and 4. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The national identity outcome is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). The villagization, national identity, cash income job and wealth index measures are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by TNPS, for columns (1) and (2), and Afrobarometer, for columns (3) to (6). Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: The Effects of *Ujamaa* on State Legitimacy and Accountability

Dep. Variable:	respect authority (1)	support one-party rule (2)	see gov. as parent (3)	trust gov. newspaper (4)	trust gov. TV/radio (5)	took action to hold gov. accountable (6)	trust own ethnic group (7)	trust other ethnic group (8)
Villagization × Treated Cohort	0.169**	0.097	0.230**	0.097	0.205***	-0.051	0.055	-0.141
	(0.071)	(0.064)	(0.111)	(0.102)	(0.069)	(0.031)	(0.097)	(0.096)
Observations	853	846	353	425	434	861	493	492
Number of clusters	52	52	47	45	45	52	45	45
R-squared	0.085	0.144	0.174	0.184	0.165	0.153	0.171	0.152
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variables are from Afrobarometer rounds 3 and/or 4 (2005-2008) and are as follows (recoded from Likert scale). (1) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: Citizens should be more active in questioning the actions of leaders. Statement 2: In our country, citizens should show more respect for authority." The dependent variable registers agreement with Statement 2.; (2) "There are many ways to govern a country. Would you disapprove or approve of the following alternatives?. Only one political party is allowed to stand for election and hold office." The dependent variable registers approval with this option.; (3) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government." The dependent variable registers agreement with Statement 1.; (4) "How much do you trust each of the following, or haven't you heard enough about them to say: Government (relative to independent) newspapers?"; (5) "Government (relative to independent) broadcasting service (TV / radio)?""; (6) "Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year" Got together with others to raise an issue"; (7) "How much do you trust each of the following types of people: "People from your own ethnic group?". (8) "People from other ethnic groups?". Questions (4), (5), (7) and (8) were asked only in Round 3, question (3) only in Round 4. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The villagization and outcome variables, except the outcome of column (6) which is binary, are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: The Effect of *Ujamaa* on Public Goods, Gender Norms, Voting

Dep. Variable:	Supports Free Schooling (1)	Agrees Women Equal Rights (2)	Agrees Women Equal Election Chances (3)	Supports CCM (4)	CCM Support in Elections (5)
Villagization × Treated Cohort	0.236***	-0.091	0.038	0.013	
	(0.049)	(0.152)	(0.104)	(0.025)	
Villagization					0.190**
					(0.077)
Observations	490	493	493	699	112
Number of clusters	45	45	45	52	56
R-squared	0.163	0.115	0.156	0.154	0.212
District FE	✓	✓	✓	✓	
Cohort FE	✓	✓	✓	✓	
Zone FE	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	
Source	Afrobarometer	Afrobarometer	Afrobarometer	Afrobarometer	Electoral data

Notes: In columns (1) to (4), the unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. These regressions are weighted using the survey weights provided by Afrobarometer. The dependent variables in columns (1) to (3) are from Afrobarometer Round 3 (2005). They are defined as follows (recoded from Likert scale). (1) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: It is better to have free schooling for our children, even if the quality of education is low. B: It is better to raise educational standards, even if we have to pay school fees." The dependent variable registers agreement with Statement A.; (2) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: In our country, women should have equal rights and receive the same treatment as men do. B: Women have always been subject to traditional laws and customs, and should remain so." The dependent variable registers agreement with Statement A.; (3) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: Women should have the same chance of being elected to political office as men. B: Men make better political leaders than women, and should be elected rather than women." The dependent variable registers agreement with Statement A. The dependent variable in column (4) is from Afrobarometer rounds 3 and 4 (2008) and is coded as a dummy for whether the respondent feels close to a particular party and this party is Chama cha Mapinduzi (CCM). In column (5), the unit of observation is a district d in year t. This column regresses support for the CCM on the district-level measure of historical villagization, controlling for the 1967 district primary school enrollment rate, CCM support in 1970, election year fixed effects, and the fixed effects indicated in the table. CCM support is defined as the product of turnout and CCM presidential vote share from the 2000 and 2005 presidential elections in Tanzania (data from Carlitz (2017)), aggregated at the level of 1967 districts, averaging over constituencies. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The villagization measure and outcomes in columns (1), (2), (3) and (5) are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

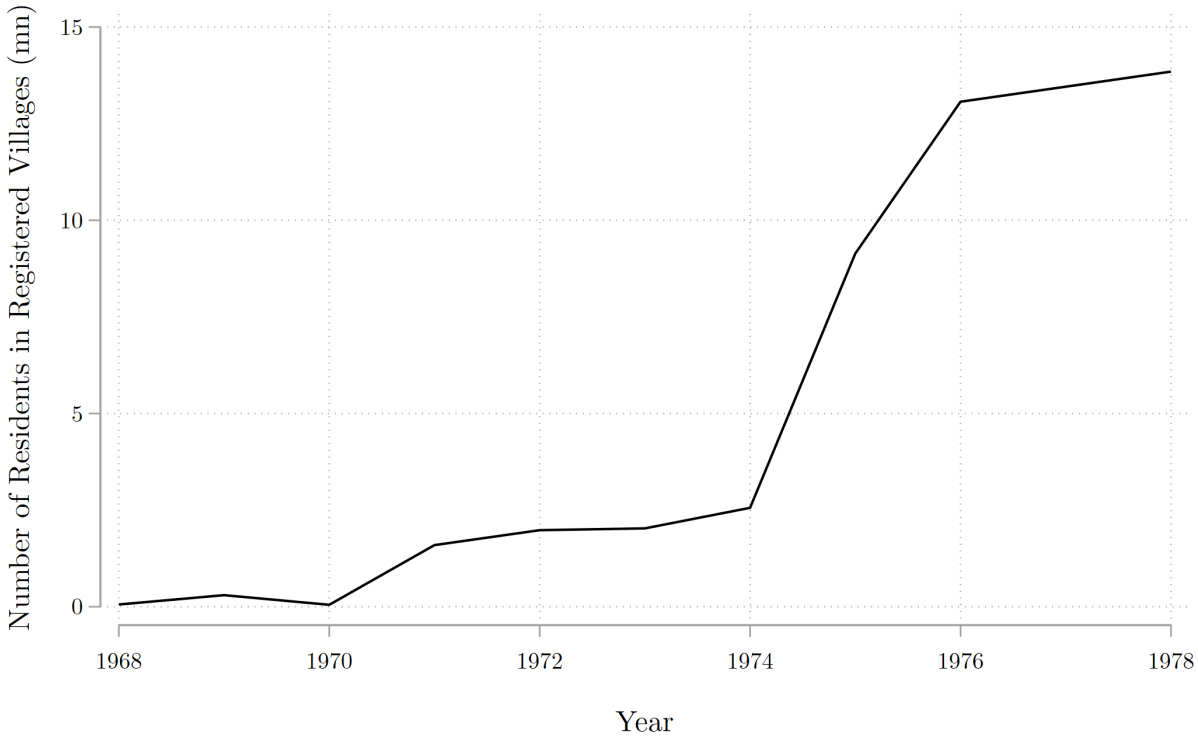
Table A.11: Correlations Between Villagization and District Characteristics in 1967

1967 district characteristic	correlation	p-value
Primary school enrollment rate in 1967 (per '000 inhabitants)	0.27	0.07
District revenue per capita in 1967 ('000 shs)	-0.03	0.78
Ethnolinguistic fractionalization in 1967	-0.09	0.34
Distance to Dar-Es-Salaam in kilometers	0.18	0.27
Centroid latitude	-0.08	0.81
Centroid longitude	0.14	0.45
Average altitude in meters	-0.19	0.41
Average slope	0.20	0.20
Hospital beds in 1967 (per '000.000 inhabitants)	0.04	0.81
Drought in 1974 (censored z-score)	-0.15	0.38
Drought in 1975 (censored z-score)	0.14	0.31
Drought in 1976 (censored z-score)	0.44	0.17
National Identity in 1967	-0.07	0.61
Share urban in 1967 (% of population)	0.03	0.84
Population density in 1967	0.22	0.27
GDP per capita in 1967 (mln shs)	0.13	0.48
Tot. market agricultural production per capita in 1967 ('000 shs)	0.23	0.29
Distance to Uganda (km)	0.51	0.33
Number of missions per capita	-0.33	0.19
Total length of roads over district area (km/km2)	0.18	0.35

Notes: This table shows pairwise correlations between villagization (share of total rural district population living in registered villages in 1978) and the variables indicated in the table, conditional on zone fixed effects. The reported p-values are robust to hetero-skedasticity. Units of observations are 1967 districts and the data includes the districts in the main sample as in Table 2. The variables and their sources are described in detail in Section 3 and Appendix A.

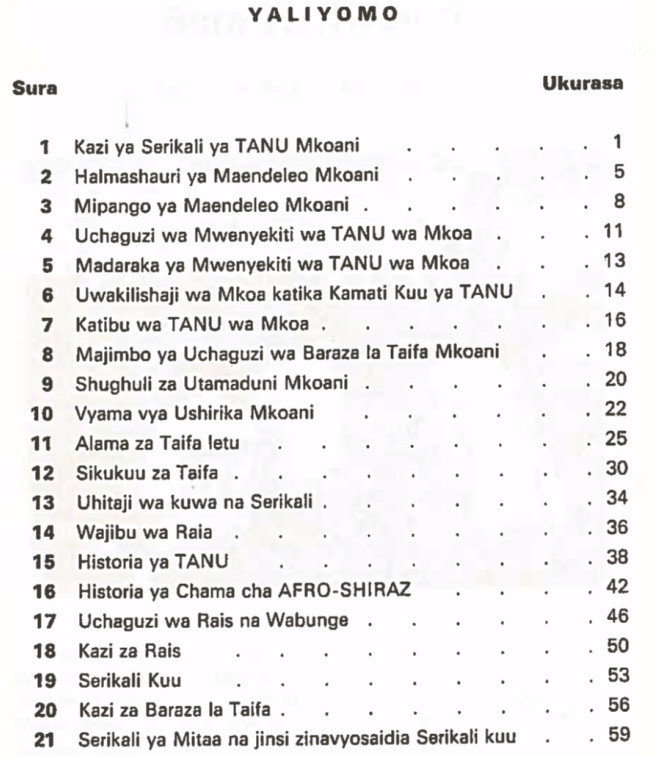
Appendix F. Additional Figures

Figure A.1: Population in Registered Villages Over Time



Notes: Authors' illustration based on data from Shao (1982). The implementation of the villagization program started after the Presidential Circular No. 1 in 1969, was accelerated in 1974 when villagization was declared mandatory, and formally ended with the repeal of the Villages and *Ujamaa* Villages Act in 1982.

Figure A.2: Political Education Textbook for 4^{th} Graders



YALIYOMO

Sura		Ukurasa
1	Kazi ya Serikali ya TANU Mkoani	1
2	Halmashauri ya Maendeleo Mkoani	5
3	Mipango ya Maendeleo Mkoani	8
4	Uchaguzi wa Mwenyekiti wa TANU wa Mkoa	11
5	Madaraka ya Mwenyekiti wa TANU wa Mkoa	13
6	Uwakilishaji wa Mkoa katika Kamati Kuu ya TANU	14
7	Katibu wa TANU wa Mkoa	16
8	Majimbo ya Uchaguzi wa Baraza la Taifa Mkoani	18
9	Shughuli za Utamaduni Mkoani	20
10	Vyama vya Ushirika Mkoani	22
11	Alama za Taifa letu	25
12	Sikukuu za Taifa	30
13	Uhitaji wa kuwa na Serikali	34
14	Wajibu wa Raia	36
15	Historia ya TANU	38
16	Historia ya Chama cha AFRO-SHIRAZ	42
17	Uchaguzi wa Rais na Wabunge	46
18	Kazi za Rais	50
19	Serikali Kuu	53
20	Kazi za Baraza la Taifa	56
21	Serikali ya Mitaa na jinsi zinavyosaidia Serikali kuu	59

CONTENTS

Notes: This image depicts the contents page of a political education textbook for 4^{th} graders. The left figure is the contents page in Swahili and the right figure is the equivalent English translation by the authors. *Sura* translates to chapter and *Ukurasa* to pages. Items underlined are the specific elements of nation-building. *Source*: Textbook titled *Elimu ya Siasa* translated as *Political Education* (United Republic of Tanzania National Department of Education, 1971a).

Figure A.3: Excerpts from Swahili Textbook

1. UMOJA NI NGUVU

Mtu anavunja kitita cha vijiti 10

(a) *"Unity is strength"*

18. UHURU NA JAMHURI

Mwakapesa tangu alipoanza shule, aliwaona watu wakisherehekea sikukuu ya "Uhuru na Jamhuri". Sherehe zilifanywa pia na watoto wa shule. Watu waliimba nyimbo zilizosifu **ukombozi** wa nchi yetu.

Mwakapesa hakuchaguliwa kuingia katika mashindano ya sikukuu mpaka alipoingia darasa la tano. Mwalimu wake alitunga nyimbo za kulaani utawala wa kigeni na za kuwasifu viongozi wa Tanzania. Mwakapesa alichaguliwa kuongoza kundi la waimbaji. Alikuwa hodari sana wa kuimbisha wenzake na wala hakuogopa kusimama mbele za watu.

Desemba tisa ilipofika, shule ya Mpuguso ilikwenda kwenye mashindano. Uwanja wa mashindano ulikuwa mbele ya Jumba la Maendeleo. Kabla ya kuimba, waimbaji walijitayarisha nyuma ya jengo ili wapate nafasi kubwa ya kutembea kigwaride.

(b) *"Independence day"*

Notes: These images depict pages from Chapter 1 (left) on *Unity is Strength* and Chapter 18 (right) on *Independence and the Republic* of the Swahili textbook *Tujifunze Lugha Yetu: Kitabu Cha Saba* translated as *Let us learn our language: Class Seven* (United Republic of Tanzania National Department of Education, 1971b).

Figure A.4: Population Census 1978, Share of Rural Population in Official Registered Villages

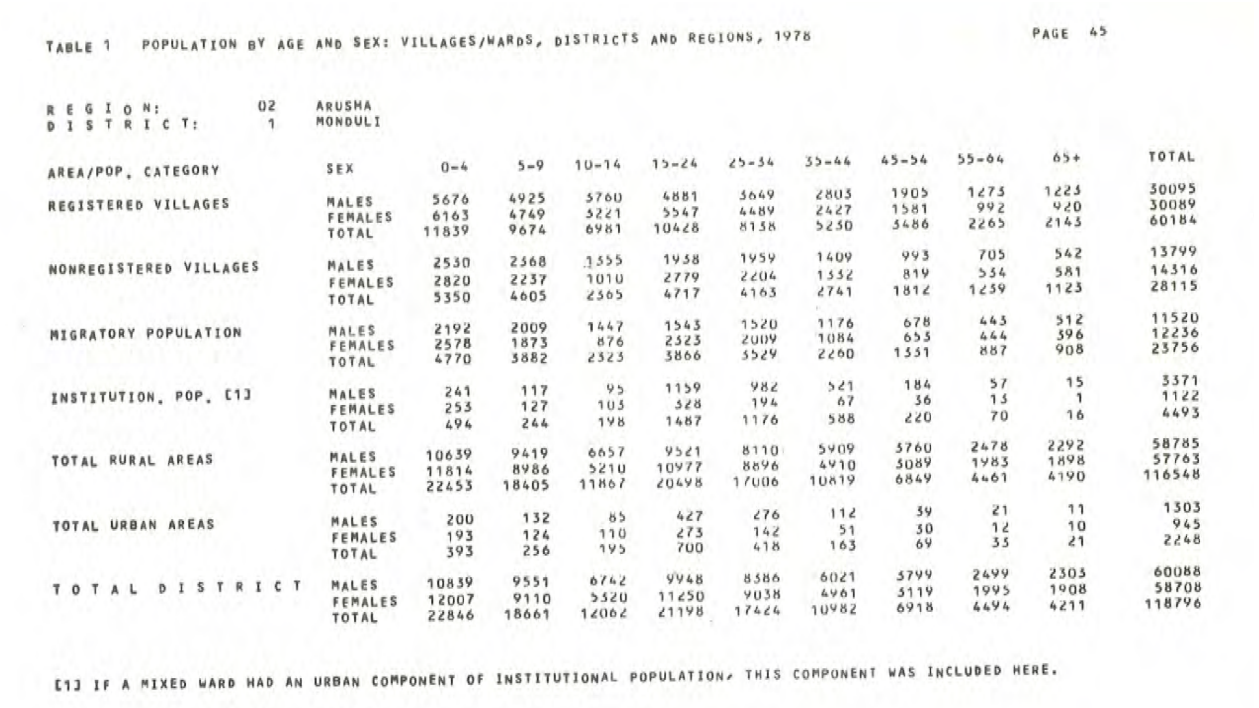


TABLE 1 POPULATION BY AGE AND SEX: VILLAGES/WARDS, DISTRICTS AND REGIONS, 1978 PAGE 45

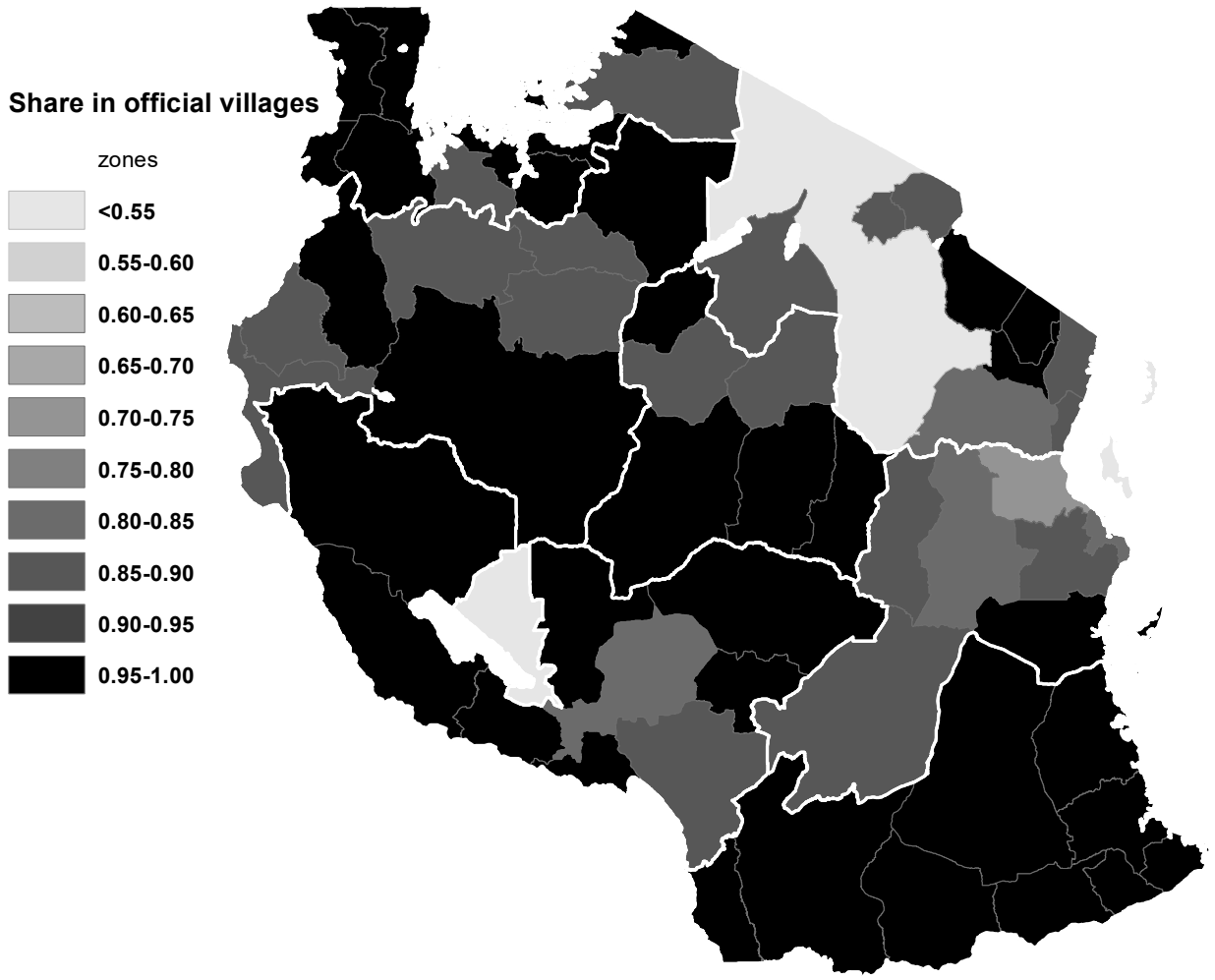
REGION: 02 ARUSHA
DISTRICT: 1 MONDULI

AREA/POP. CATEGORY	SEX	0-4	5-9	10-14	15-24	25-34	35-44	45-54	55-64	65+	TOTAL
REGISTERED VILLAGES	MALES	5676	4925	5760	4881	3649	2803	1905	1273	1223	30095
	FEMALES	6163	4749	5221	5547	4489	2427	1581	992	920	30089
	TOTAL	11839	9674	6981	10428	8138	5230	3486	2265	2143	60184
NONREGISTERED VILLAGES	MALES	2530	2368	1355	1938	1959	1409	993	705	542	13799
	FEMALES	2820	2237	1010	2779	2204	1332	819	534	581	14316
	TOTAL	5350	4605	2365	4717	4163	2741	1812	1239	1123	28115
MIGRATORY POPULATION	MALES	2192	2009	1447	1543	1520	1176	678	443	512	11520
	FEMALES	2578	1873	876	2323	2009	1084	653	444	396	12236
	TOTAL	4770	3882	2323	3866	3529	2260	1331	887	908	23756
INSTITUTION, POP. [1]	MALES	241	117	95	1159	982	521	184	57	15	3371
	FEMALES	253	127	103	328	194	67	36	13	1	1122
	TOTAL	494	244	198	1487	1176	588	220	70	16	4493
TOTAL RURAL AREAS	MALES	10639	9419	6657	9521	8110	5909	3760	2478	2292	58785
	FEMALES	11814	8986	5210	10977	8896	4910	3089	1983	1898	57763
	TOTAL	22453	18405	11867	20498	17006	10819	6849	4461	4190	116548
TOTAL URBAN AREAS	MALES	200	132	85	427	276	112	39	21	11	1303
	FEMALES	193	124	110	273	142	51	30	12	10	945
	TOTAL	393	256	195	700	418	163	69	33	21	2248
TOTAL DISTRICT	MALES	10839	9551	6742	9948	8386	6021	3799	2499	2303	60088
	FEMALES	12007	9110	5320	11250	9038	4961	3119	1995	1908	58708
	TOTAL	22846	18661	12062	21198	17424	10982	6918	4494	4211	118796

[1] IF A MIXED WARD HAD AN URBAN COMPONENT OF INSTITUTIONAL POPULATION, THIS COMPONENT WAS INCLUDED HERE.

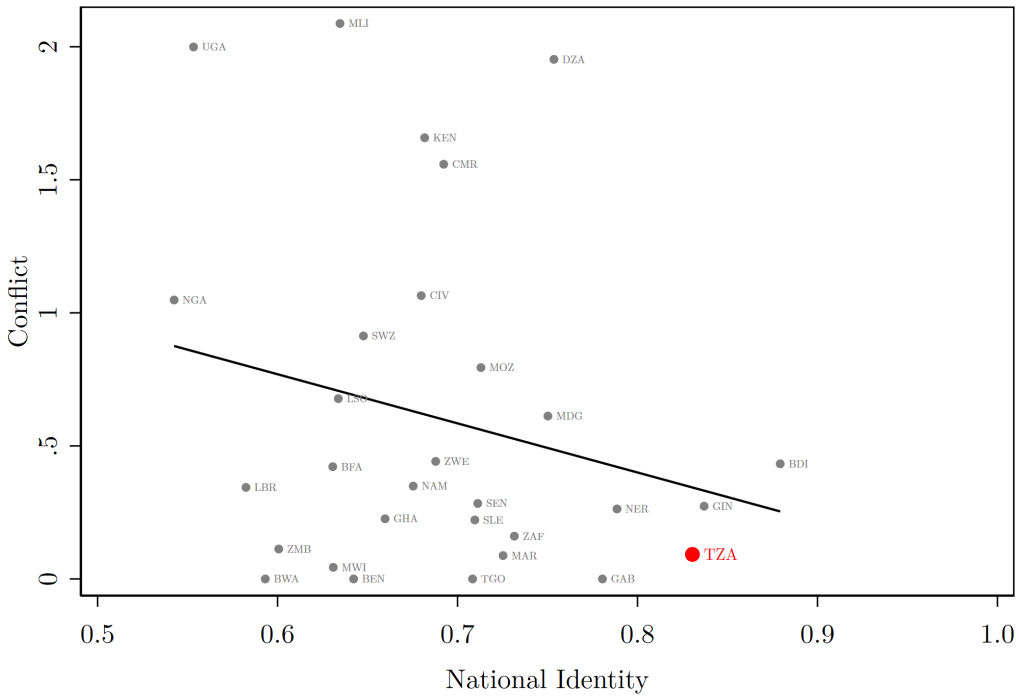
Notes: This image shows an excerpt from the 1978 Tanzania Population Census (Bureau of Statistics, Ministry of Planning and Economic Affairs, 1981), retrieved from the Herskovits Library of African Studies at Northwestern University.

Figure A.5: Distribution of Villagization



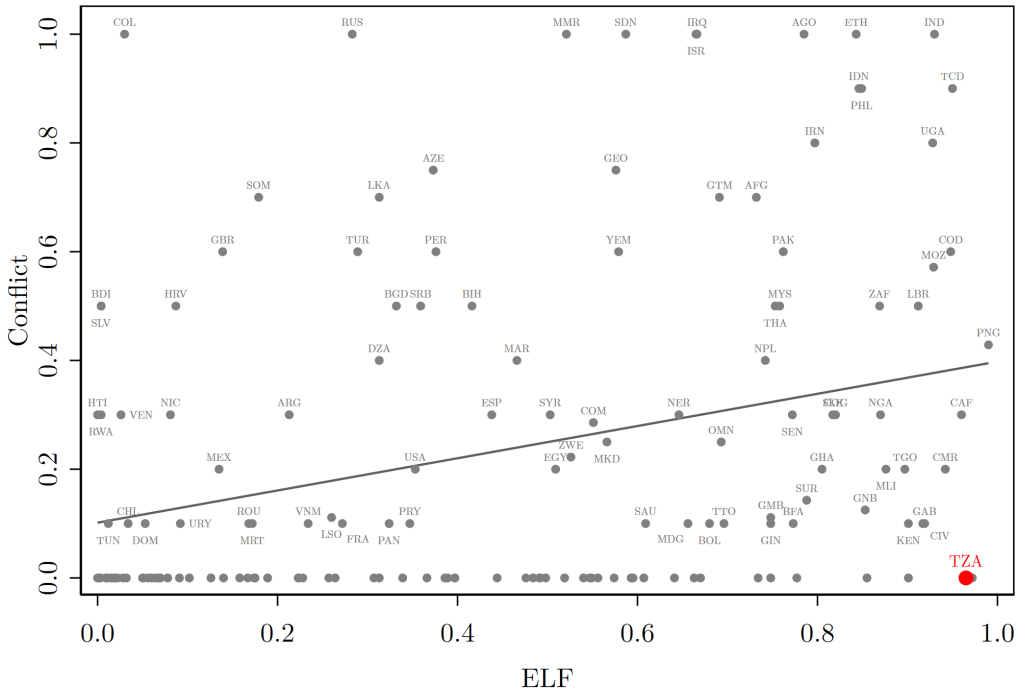
Notes: This figure depicts our district-level measure of villagization: the share of the rural population living in official registered villages according to the 1978 Tanzania Population Census (at 1967 district borders). The population not living in official registered villages includes those in unregistered villages and those outside of villages (scattered and migratory population). Thick white lines denote zone boundaries. Data for Songwe district is missing. Dar es Salaam (Mzizima) and islands (Zanzibar, Pemba, Mafia) are excluded from the sample. Sources: Authors' illustration. Base map from The Humanitarian Data Exchange (HDX) by United Nations OCHA (https://data.humdata.org/, accessed July, 2021).

Figure A.6: National Identity and Conflict in Africa



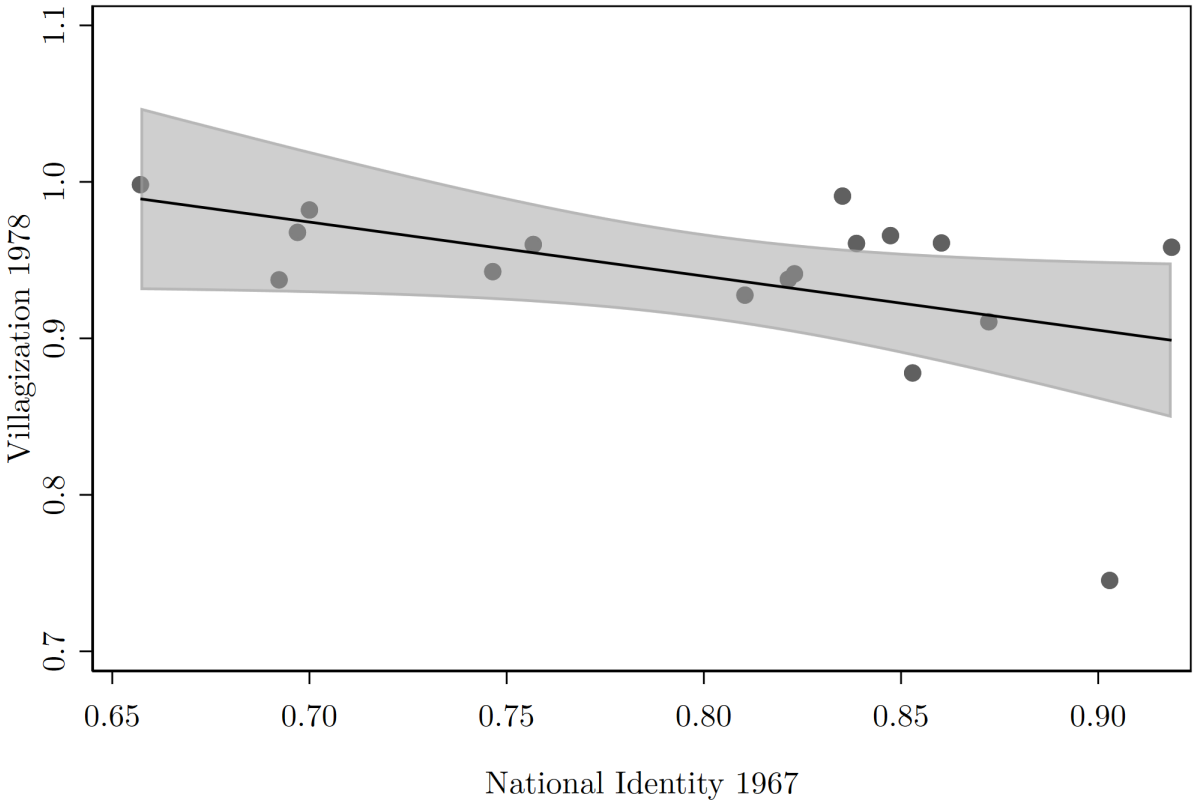
Notes: This plot illustrates the correlation between the mean number of battles per 1 million inhabitants (defined as "conflict") and average expressed feelings of national identity (as opposed to ethnic identity) for 33 African countries between 2005 and 2014. Conflict data is from ACLED (Raleigh et al., 2010) and population data is from the World Development Indicators (World Bank, 2020). Country average of national identity from Afrobarometer Rounds 3-6. Data is at the country level.

Figure A.7: Conflict and Ethnolinguistic Fractionalization



Notes: This plot illustrates the correlation between conflict (incidence of internal armed conflict with at least 25 deaths in the last 5 years based on data from Gleditsch et al. (2002)) and ethnolinguistic fractionalization (index based on data from Fearon (2003)) at the country level between 1960 and 2008. The data are taken from Esteban et al. (2012), which provides more details on the data.

Figure A.8: National Identity in 1967 and Villagization in 1978



Notes: This plot shows the correlation (with 95% confidence interval) between region-level villagization in 1978 and average feelings of national identity in 1967. The estimated coefficient is -0.35 and the p-value is 0.10 (standard errors are robust to hetero-skedasticity). Villagization is defined as the fraction of the rural region population living in official registered villages according to the 1978 Tanzania Population Census. Region-level villagization is computed as the within-region average of 1967 district level villagization. The data on national identity is based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 by Prewitt et al. (1970). The original variable is coded as 1 if a respondent feels that the nation is more important than the tribe and 0 otherwise.

Online Appendix References

Barker, J. S. (1974). "ujamaa" in cash-crop areas of tanzania: Some problems and reflections. *Journal of African Studies 1*(4), 441.

Boesen, J., B. S. Madsen, and T. Moody (1977). Ujamaa–socialism from above. 181 pp. scandinavian institute of african. *Studies, Uppsala*.

Bureau of Statistics, Ministry of Planning and Economic Affairs (1981). *1978 Population Census*. Dar es Salaam: United Republic of Tanzania.

Carlitz, R. D. (2017). Money flows, water trickles: Understanding patterns of decentralized water provision in tanzania. *World Development 93*, 16–30.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Ergas, Z. (1980). Why did the Ujamaa village policy fail? Towards a global analysis. *The Journal of Modern African Studies 18*(3), 387–410.

Esteban, J., L. Mayoral, and D. Ray (2012). Ethnicity and conflict: An empirical study. *American Economic Review 102*(4), 1310–42.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of economic growth 8*, 195–222.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed conflict 1946-2001: A new dataset. *Journal of peace research 39*(5), 615–637.

Hayuma, A. (1980). Dodoma: The planning and building of the new capital city of tanzania. *Habitat International 5*(5-6), 653–680.

Hedde-von Westernhagen, C. and B. Becker (2022). Mapping missions: New data for the study of african history: Social and economic history. *Research Data Journal for the Humanities and Social Sciences 7*(1), 1–33.

Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.

Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press.

Jedwab, R. and A. Storeygard (2022, 06). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association 20*(1), 1–38.

Jensen (1968). *Government of Tanzania District Statistics*. Dar es Salaam: Government Printers.

McHenry, D. E. (1979). *Tanzania's Ujamaa Villages. The Implementation of a Rural Development Strategy*. Institute of International Studies, University of California, Berkeley.

Prewitt, K., G. Von Der Muhll, and D. Court (1970). School experiences and political socialization: A study of tanzanian secondary school students. *Comparative Political Studies 3*(2), 203–225.

Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing ACLED: an armed conflict location and event dataset: Special data feature. *Journal of Peace Research 47*(5),

651–660.

Roberts, G. (2014). The uganda–tanzania war, the fall of idi amin, and the failure of african diplomacy, 1978–1979. *Journal of Eastern African Studies 8*(4), 692–709.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Shao, I. F. (1982). A neo-colony and its problems during the process of attempting to bring about socialist rural transformation: the case of Tanzania. In *Taamuli: A Political Science Forum*, Volume 12, pp. 29–46.

Sitari, T. (1983). Settlement changes in the bagamoyo district of tanzania as a consequence of villagization. *Fennia-International Journal of Geography 161*(1), 1–90.

Stabler, E. (1979). Kenya and tanzania: Strategies and realities in education and development. *African Affairs 78*(310), 33–56.

United Republic of Tanzania National Department of Education (1971a). *Elimu ya Siasa 6: Taifa na Serikali Wetu*. Tanzania Publishing House.

United Republic of Tanzania National Department of Education (1971b). *Tujifunze lugha yetu : kitabu cha saba*. Tanzania Publishing House.

World Bank (2020). *World Development Indicators 2020*. The World Bank.